DOCUMENT RESUME

ED 437 941 IR 057 532

AUTHOR	Davis-Tanous, Jennifer R.
TITLE	XML: A Language To Manage the World Wide Web. ERIC Digest.
INSTITUTION	ERIC Clearinghouse on Information and Technology, Syracuse, NY.
SPONS AGENCY	Department of Education, Washington, DC.
REPORT NO	EDO-IR-1999-10
PUB DATE	1999-12-00
NOTE	4p.
CONTRACT	ED-99-CO-0005
AVAILABLE FROM	ERIC Clearinghouse on Information & Technology, Syracuse University, 621 Skytop Rd., Suite 160, Syracuse, NY 13244-5290; Tel: 315-443-3640; Tel: 800-464-9107 (Toll Free); Fax: 315-443-5448; e-mail: eric@ericir.syr.edu; Web site: <http://ericir.syr.edu/ithome>.
PUB TYPE	ERIC Publications (071) -- ERIC Digests in Full Text (073)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Hypermedia; *Programming Languages; *Screen Design (Computers); *World Wide Web
IDENTIFIERS	ERIC Digests; Gateway to Educational Materials; HTML; Metadata; SGML; Web Browsers; *Web Page Design; *XML

ABSTRACT

This digest provides an overview of XML (Extensible Markup Language), a markup language used to construct World Wide Web pages. Topics addressed include: (1) definition of a markup language, including comparison of XML with SGML (Standard Generalized Markup Language) and HTML (HyperText Markup Language); (2) how XML works, including sample tags, metadata, and use of XML by the GEM (Gateway to Educational Materials) Project; (3) customizing XML for individual needs, including the DTD (Document Type Definition), i.e., a formal set of markup elements developed as a standard for professionals in a particular field; and (4) making XML work on the Web, including limitations of older browsers. (Contains 12 references.) (MES)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 437 941

XML: A Language to Manage the World Wide Web

ERIC Digest

Jennifer R. Davis-Tanous

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it.
- ☐ Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

IR057532

Clearinghouse on Information & Technology
IT
621 Skytop Road, Suite 160
Syracuse University
Syracuse, New York 13244-5290

ERIC®
DIGEST

December 1999 EDO-IR-1999-10

XML: A Language to Manage the World Wide Web

by Jennifer R. Davis-Tanous

Introduction

Extensible Markup Language, or XML, is poised to become the standard markup language used to construct Web pages on the World Wide Web. Extensible Markup Language incorporates components of both Standard Generalized Markup Language (SGML) and HyperText Markup Language (HTML), resulting in a flexible language that is user-friendly and supports many different applications.

First, it is essential to understand how a computer "reads" a Web page. HyperText Markup Language displays pages on the World Wide Web by tagging different elements in a document (Webopedia, 1999). Because HTML is a very basic language, the demand for formatting data rather than just displaying it has surpassed HTML's capabilities. Therefore, XML has been introduced as a possible solution to the increased demand for formatted information on Web pages. XML provides a standard for Web authors that can be read by different browsers and different computer platforms. Extensible Markup Language seeks to do away with vendor-specific markup language (compatible with only Internet Explorer or Netscape Navigator, for example). Extensible Markup Language will make the Web a more efficient education tool because it will allow for more accurate searching. The data in XML Web pages will be structured and not just displayed.

What Is a Markup Language?

Simply stated, a Web page must be written in a markup language for a computer Web browser to interpret how to display that page. Standard Generalized Markup Language (SGML) is a complex language that allows a programmer to format documents. HyperText Markup Language is a language described in SGML, and widely regarded as the standard for Web publishing. HyperText Markup Language is quite austere compared to SGML, and therefore limited. HyperText Markup Language uses tags to describe how data will be presented on a Web page. For instance, the tag element <BOLD> is used to make text appear in boldface (Bosak, 1999). Of course, the Web is a dynamic environment, and new demands are made of HTML all the time. As more elements are added to HTML, problems arise with browser compatibility. Something that works well in Netscape Navigator might fail miserably in Internet Explorer.

Also, HTML can make an attractive Web page fairly easily, but it doesn't tell the computer a thing about content. With websites proliferating at an astounding rate, the need presents itself for a markup language that is both multi-browser compatible and capable of formatting data so that information on the World Wide Web is found more quickly and easily. Therefore, XML was developed. Because XML is not as pared down as HTML, it can use the complexity of SGML to make Web pages more active. The result will be a faster World Wide Web, with more reliable search results.

How XML Works

Extensible Markup Language allows a person to invent an array of tags to describe their text document (Bray, 1997). In HTML, there are a limited number of tags, such as <BOLD> or <ITALIC>, and these tags format text—that's it. In XML, a person could invent a set of tags to describe, for instance, a lesson plan. Such a set of tags might look something like this:

<LESSON PLAN>
<SUBJECT>English Literature</SUBJECT>
<TITLE>An Introduction to Shakespeare</TITLE>
<CONCEPTS>
<P>The main concepts covered in this lesson are the life of William Shakespeare (i.e., his childhood, early acting career, life as a playwright, his personal life) and the Elizabethan Era. </P>
</CONCEPTS>
</LESSON PLAN>

If an English teacher wanted to mine the Web for lesson plans, XML would allow search engines to conduct a much more productive search based on the tags used, similar to those illustrated above.

Suppose an educator was interested in developing a lesson plan on the life of William Shakespeare. Entering the words "William Shakespeare" in a typical search engine now could result in thousands and thousands of hits, with relatively few of educational value. With XML, search engines will search both the tags and the content of the page, thus bringing up "Lesson Plan" or "Literature," and winnowing the search results to the rich, relevant data needed. This type of tagging is referred to as metadata, or literally, "data about data." In the same way, it would be much easier to find information about the movie *Shakespeare In Love*, because the metatag for that site would be <MOVIE>, or something similarly descriptive.

The Gateway to Educational Materials (GEM) Project is an online ERIC resource for Internet based lesson plans and curriculum units. GEM will be able to build a set of XML tags which specify exactly how the Web pages for these educational materials should be put together. As a result, a standard will be developed, not only in how the pages appear to the user, but in how the search engines interpret the data that they contain. HyperText Markup Language will give the GEM project a distinctive "look" via images, colors, and fonts. More importantly, XML will create a standard for how the GEM information is formatted, much as described previously with the lesson plans. The key concept here is the containment of data. Being able to find the data on a Web site in an organized fashion greatly increases the value of that Web site—and XML can do this.

Customizing XML for Individual Needs

Now, chaos could easily erupt if everyone in charge of a website decided to arbitrarily design his or her own set of metatags as descriptors. However, the potential for specific groups of people, such as educators or those at the GEM project, to customize their own particular sets of elements is enormous. When a set of metatags is developed for a particular interest group, it is referred to as a Document Type Definition (DTD). By fashioning a DTD, a formal set of markup elements can be developed as a standard for professionals in a particular field. The DTD names the elements and defines what, where, and how they may be used (Flynn, 1999). The DTD will also tell the author what tags are acceptable, how the tags must be arranged within each other, and in what order they need to appear. The process is similar to preparing a composition paper. A teacher giving a writing assignment would expect that the paper's introduction would come first, then the body, followed by the conclusion. She would expect students to place clauses inside sentences, and sentences inside paragraphs. The students would be required to use this DTD, but they could fill in their own "data." If this were a DTD for a history class, then the content of the paper would have something to do with history.

The DTD holds many implications for streamlining data that come from many resources but relate to a particular thing. The student information form, which college freshmen fill out as they enter college for the first time, could be completed using a form on various websites using the same DTD. Because all the data would be housed exactly the same way, it could be much more easily mined for important information about this group of students. Instead of having to manipulate huge sets of raw data, researchers would find the data already organized in a predetermined way.

Making It Work on the World Wide Web

Extensible Markup Language is still being adapted to the limitations of browsers. Originally, XML required the use of Cascading Style Sheets (CSS). In essence, CSS allows Web authors to write their own markup language to determine how the content of a particular page will be displayed. The Web author can write a piece of markup code, for instance: H2 {font: 24pt Helvetica; font-weight: bold;}. The code is contained within the stylesheet, which means that every time the author uses the HTML H2 in the body of the page, it will automatically be 24pt bold Helvetica. By using CSS, the Web author needs to define his or her expectations for H2 only once (within the stylesheet) instead of every time it occurs in the body of the Web page.

Unfortunately, CSS commonly fails with today's browsers. A stylesheet that works for Navigator might not work in Explorer, and vice versa. A font, such as Helvetica or Arial, might be specific to only one of the browsers. This might seriously impact the appearance of the Web page to any users on another browser. Moreover, older versions of browsers will not be able to handle CSS, so it is important for Web authors to consider how many of their potential users will be on older browsers.

Summary

For anyone who has ever dabbled in Web authoring, the reassuring news is that XML promises to be just as easy to learn as HTML. The biggest change is that the Web author must write or borrow a DTD before beginning. As XML becomes more pervasive, expect to find DTDs readily available in a variety of subject matters. Cascading Style Sheets are also relatively simple to learn and use, and pages use less bandwidth because specifics about certain tags are contained within the stylesheet instead of throughout the body of the Web page. Because XML is still a relatively new development, browsers are not yet being marketed as XML-compatible. HTML and SGML documents will still be viewable while browsers begin to implement XML (Flynn, 1999). Extensible Markup Language holds great promise for organizing data on the World Wide Web. Its capacity for formatting data will be a great leap forward for all those who are connected to the Internet, either as Web authors or Web users. *

Bibliography and Further Reading

Beginning XML. (1998). The Mining Company. <*http://html.miningco.com/msubXMLintro.htm*> (version current at 03 Dec 1999).

Bosak, J. & Bray, T. (1999). XML and the second-generation. *Scientific American*, 89, 93.

Bray, T.(September, 1997). Beyond HTML: XML and automated web processing. Internet WWW page, at URL: <*http://developer.netscape.com/viewsource/bray_xml.html*> (version current at 03 Dec 1999).

Flynn, P. (June, 1999). Frequently asked questions about the extensible markup language, Version 1.5. Internet WWW page, at URL: <*http://www.ucc.ie/xml/#FAQ-DOCTYPE*> (version current at 03 Dec 1999).

Gateway to Educational Materials Project. Internet WWW page, at URL: <*http://www.thegateway.org*> (version current at 03 Dec 1999).

Hockey, S. (1997). *Making technology work for scholarship: Investing in the data.* Paper presented at the Conference on Scholarly Communication and Technology (Atlanta, GA, April 24-25, 1997). (ED 414 932)

Lander, R. (1998). A tutorial in XML and XSL authoring. Internet WWW page, at URL: <*http://pdbeam.uwaterloo.ca/~rlander/XML_Tutorial/index.html*> (version current at 03 Dec 1999).

A leaner, meaner markup language. (June, 1997). *Online & CD-ROM Review, 21*(3), 181-84. (EJ 547 847)

Lewis, J. D. (1998). XML: An introduction. *OCLC Systems & Services, 14*(1), 51-52. (EJ 566 526)

Webopedia. (1999). Internet.Com Corp. <*http://webopedia.internet.com/*> (version current at 03 Dec 1999).

World Wide Web Consortium. (1997). Extensible markup language. Internet WWW page, at URL: <*http://www.w3.org/XML/*> (version current at 03 Dec 1999).

XML.Com (1998). Internet WWW page, at URL: <*http://www.xml.com/xml/pub*> (version current at 03 Dec 1999).

This Digest was prepared by Jennifer R. Davis-Tanous <*jdavis@abacus.bates.edu*>, Career Information and Technology Coordinator at Bates College, and a graduate student in the ISDP-MLS program at Syracuse University.

ERIC Digests are in the public domain and may be freely reproduced and disseminated.

ERIC Clearinghouse on Information & Technology, Syracuse University, 621 Skytop Road, Suite 160, Syracuse, New York 13244-5290; (315) 443-3640; (800) 464-9107; Fax: (315) 443-5448; E-mail: *eric@ericir.syr.edu*; URL: *http://ericir.syr.edu/ithome*

This publication is funded in part with Federal funds from the U. S. Department of Education under contract number ED-99-CO-0005. The content of this publication does not necessarily reflect the views or policies of the U. S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U. S. government. The U.S. Department of Education's web address is: http://www.ed.gov/